Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Environmental Sciences 8 (2011) 730 – 736

Procedia Environmental Sciences

ICESB 2011: 25-26 November 2011, Maldives

Analysis of Codon Usage Patterns of Tianfu Goose Interferon Alpha

Fei Liu*, Liqin Gao, Anchun Cheng, Mingshu Wang, Qinglin Han

Avian Disease Research Center, College of Veterinary Medicine of Sichuan Agricultural University Ya'an, Sichuan, 625014, China

Abstract

Interferons(IFNs) are a large family of cytokines, belonging to glycoproteins. As for the antiviral and antineoplastic activity of IFNs, IFNs were widely studied. Based of recent gene sequence analyses, we found that codon usages were variation in different IFN genes. Codon usage in a sample of 5 genes from the interferon alpha has been analysed using multivariate statistical analysis. Goose and duck IFN-α gene shared the similar patterns, and differ from chicken IFN-α gene codon usage biases pattern. In other words, goose and duck had a much closer relationship than chicken. Asn(L), His(H), Ala(A) and Pro(P) were the preferred amino acids. RSCU values of 25 codons were greater than 1.0, and these codons had a much higher frequency of occurrence. In the accordance with the comparison, E.coli was much suitable for the expression of Tianfu goose IFN-α gene than Yeast and Human system.

Selection and/or peer_review under responsibility of the Asia Pacific Chemical Biological & Environmental Engineering Society (APCBEES)

Keywords: Tianfu goose IFN-α, codon usage, analysis, comparison.

1. Introduction

Codons, a tri-nucleotide sequence constituting the base coding unit, encode amine acid. Synonymous codons, encoding the same amine acid, are not equally used in genes, which is called codon usage bias. The frequencies of synonymous codons used distinctly among genomes[1, 2]. It was reported that the biased usage of synonymous codons was related to multiplicitas factors, such as gene length[3, 4], GC content[5], recombination rate, gene expression lever[6], and density of genes[7]. The shaping factors of synonymous codon are discriminatory among different organisms. To investigate the codon usage

* Corresponding author. Tel.: 13981614050
E-mail address: liufei@sicau.edu.cn.

Selection and/or peer-review under responsibility of the Asia-Pacific Chemical, Biological & Environmental Engineering Society (APCBEES)
doi:10.1016/j.proenv.2011.10.111

patterns, many codon usage indices were calculated, such as the codon adaptation index (CAI), effective number of codons used in a gene (ENC), relative synonymous codon usage (RSCU), G+C content at eh third positions of codons (GC3s) and so on.

Interferons(IFNs) are a large family of cytokines, which belong to glycoproteins. As for the antiviral and antineoplastic activity of IFNs, IFNs were widely studied.

Recently, IFN-α of three goose breeds was amplified in China, such as Dongbei White goose IFN-α[8], Shitou goose IFN-α[9], and Tianfu goose IFN-α. And human IFN-α was widely used in clinical medicine, to treat hairy cell leukemia[10], chronic myelogenous leukemia, B and T cell lymphomas, melanomas, and Kaposi's sarcoma[11, 12].Though the IFN of poultry was first gained, the evolution of poultry IFN was hysteretic, especially the evolution of goose IFN.

Based on Tianfu goose IFN-α gene, the synonymous codon usage pattern was analyzed. In this study, we performed a correlation analysis of codon bias of goose IFN-α.genes and duck IFN-α genes. Meanwhile, were analysied the rare codons of the Tianfu goose IFN-α ORF. Moreover, the codon usage patterns in the Tianfu goose IFN-α were compared with E.coli, yeast and Human. These information could not only improve the understanding of the codon usage patterns factors, but also provide more deep cognitions of goose IFN-α.

2. Materials and methods

2.1. IFN-α and gene sequences

Tianfu goose gene was obtained from Key Laboratory of Animal Disease and Human Health of Sichuan Agriculture University, and the other reference IFN-α gene nucleotide sequences were obtained from GenBank.

2.2. Codon usage bias analysis in the IFN-α genes

The codon usage indices, CAI(Codon Adaptation Index)[13], Fop (Frequency of Optimal codons)[14], CBI (Codon Bias Index)[15], ENC (the effective number of codons)[16], RSCU (relative synonymous codon usage), GC3s (G+C content 3rd position of synonymous codons), GC content were calculated by the programs of the codonw Correspondence Analysis of Codon Usage on line. Meanwhile, data on codon usage of Tianfu goose IFN-α was obtained from the codonw.

2.3. Analyze the optimum expression host of Tianfu goose IFN-α gene among 3 other argnisms

According to the definition of optimum expression host, the comparison of codon usage bias among Tianfu goose IFN-α gene with those of E.coli, yeast and human was performed.

3. Results

3.1. Analyze the codon feature of Tianfu goose IFN-α

The results obtained by codonw on line analysis of codon number, fraction, frequency and RSCU (not shown). According to the information, histogram was structured to show the amino acid number (Figure 1) and the RSCU of codon (Figure 2). It was obvious that the number of Asn(L), His(H), Ala(A) and Pro(P) was much greater than others. RSCU was defined as the proportion of observed frequency of codons to the expeted frequency[5, 13]. And if RSCU value was higher than 1.0, it indicated that the corresponding

codon was used more often than expected, nevertheless the adverse is true for RSCU values lower than 1.0. From Figure 2, we learned that RSCU values of 25 codons were greater than 1.0, such as GGG, TGC, GAC, GAG, TTC, GGC, CTC, GTA, and so on.

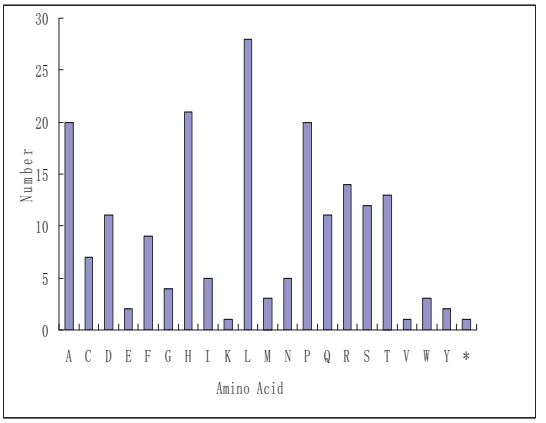

Figure 1 The number of Amino Acid (Tianfu goose IFN-α)

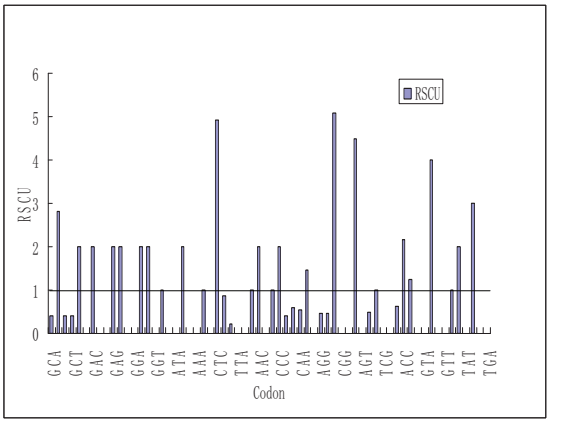

Figure 2 RSCU values of codons

3.2. Analyze the codon usage patterns in IFN-α genes

CAI is a dimension of the relative adaptiveness of the codon usage of a gene for the codon usage of highly expressed genes[13]. Codon bias index is another dimension of directional codon bias, it estimates the extent to which a gene uses a subset of optimal codons[15]. Frequency of Optimal codons**,** Fop, is the ratio of optimal codons to synonymous codons[17]. The effective number of codons of a gene(ENC) was used to quantify the codon usage bias of a gene[18]. The average value of CAI, CBI, Fop, ENC, GC content and GC3s was 0.284, 0.339, 0.608, 31.80, 66.21% and 85.71% respectively. It was found that those values of IFN-α genes of goose and duck all lower than the average value, whereas chicken IFN-αs had much higher values than the average value. It can be concluded that goose and duck IFN-α gene shared similar codon usage bias pattern, and differed from chicken IFN-α gene.

Table 1 Summary analysis of 7 IFN-αgenes

GenBank accession no.	Animals' IFN-α	CAI	CBI	Fop	ENC	GC(%)	GC3s (%)
HQ115583	Tianfu goose IFN-α	0.268	0.310	0.589	30.49	68.60%	89.20%
AY524422	Northeast goose IFN-α	0.269	0.309	0.589	30.45	67.50%	87.60%
EU029159	Shitou goose IFN-α	0.262	0.314	0.591	30.60	69.30%	89.80%
AY879230	Sichuan Ma duck IFN-α	0.279	0.327	0.600	30.41	67.40%	87.60%
AB128861	Beijing duck IFN-α	0.275	0.321	0.595	29.95	67.50%	87.00%
HQ008781	White Leghorn chicken IFN-α	0.318	0.396	0.647	35.57	61.50%	79.10%
EU334503	Beijing fatty chicken IFN-α	0.315	0.396	0.647	35.19	61.70%	79.70%

3.3. Analyze the optimum expression host of Tianfu goose IFN-α gene among 3 other argnisms

Information of comparison analysis(Table 2) revealed that there were 47 codons which showed distinct usage differences between Tianfu goose IFN-α to E.coli, and 46 codons and 45 codons showing distinct usage difference between Tianfu goose IFN-α to Yeast and Human, respectively. It was that the E.coli system was much suitable for the gene expression.

Table 2 Comparison of codon preferences between the DuIFN-αlpha gene and E. coli, yeast and human

Amino acid	**Codon**	**1/1000**				**GoIFN-α/**		
		IFN-α	*E. coli*	*Yeast*	*Human*	*E. coli*	*Yeast*	*Human*
A(Ala)	GCA	10.417	20.600	16.100	16.100	0.506	0.647	0.647
A	GCC	72.917	25.500	12.500	28.400	2.859	5.833	2.568
A	GCG	10.417	31.700	6.100	7.500	0.329	1.708	1.389
A	GCT	10.417	15.600	21.100	18.600	0.668	0.494	0.560
C(Cys)	TGC	36.458	6.900	4.700	12.200	5.284	7.757	2.988
C	TGT	0.000	5.500	8.000	10.000	0.000	0.000	0.000
D(Asp)	GAC	57.292	18.600	20.200	25.600	3.080	2.836	2.238
D	GAT	0.000	32.100	37.800	21.900	0.000	0.000	0.000
E(Glu)	GAA	0.000	38.200	48.500	29.000	0.000	0.000	0.000
E	GAG	10.417	17.700	19.100	39.900	0.589	0.545	0.261
F(Phe)	TTC	46.875	16.900	18.200	20.600	2.774	2.576	2.275
F	TTT	0.000	23.200	26.100	17.100	0.000	0.000	0.000
G(Gly)	GGA	0.000	9.000	10.900	16.400	0.000	0.000	0.000
G	GGC	10.417	27.900	9.700	22.500	0.373	1.074	0.463
G	GGG	10.417	11.300	6.000	16.300	0.922	1.736	0.639
G	GGT	0.000	24.400	24.000	10.800	0.000	0.000	0.000
H(His)	CAC	109.375	9.800	7.700	15.000	11.161	14.205	7.292
H	CAT	0.000	13.600	13.700	10.500	0.000	0.000	0.000
I(Ile)	ATA	0.000	5.400	17.800	7.700	0.000	0.000	0.000
I	ATC	26.042	24.200	17.000	21.600	1.076	1.532	1.206
I	ATT	0.000	29.800	30.400	16.100	0.000	0.000	0.000
K(Lys)	AAA	0.000	33.200	42.200	24.100	0.000	0.000	0.000
K	AAG	5.208	10.700	30.700	32.200	0.487	0.170	0.162
L(Leu)	CTA	0.000	4.000	13.300	7.800	0.000	0.000	0.000
L	CTC	119.792	11.000	5.400	19.800	10.890	22.184	6.050
L	CTG	20.833	50.900	10.400	39.800	0.409	2.003	0.523
L	CTT	5.208	11.700	12.100	13.000	0.445	0.430	0.401
L	TTA	0.000	13.900	26.700	7.500	0.000	0.000	0.000
L	TTG	0.000	14.000	27.000	12.600	0.000	0.000	0.000

M(Met)	ATG	15.625	27.000	20.900	22.200	0.579	0.748	0.704
N(Asn)	AAC	26.042	21.400	24.900	19.500	1.217	1.046	1.335
N	AAT	0.000	18.600	36.300	16.700	0.000	0.000	0.000
P(Pro)	CCA	26.042	8.500	18.200	16.700	3.064	1.431	1.559
P	CCC	52.083	5.800	6.800	20.100	8.980	7.659	2.591
P	CCG	10.417	21.800	5.300	6.900	0.478	1.965	1.510
P	CCT	15.625	7.300	13.600	17.300	2.140	1.149	0.903
Q(Gln)	CAA	15.625	15.000	27.500	12.000	1.042	0.568	1.302
Q	CAG	41.667	29.500	12.100	34.100	1.412	3.444	1.222
R(Arg)	AGA	0.000	2.900	21.300	11.500	0.000	0.000	0.000
R	AGG	5.208	1.900	9.200	11.400	2.741	0.566	0.457
R	CGA	5.208	3.900	3.000	6.300	1.335	1.736	0.827
R	CGC	57.292	21.000	2.600	10.700	2.728	22.035	5.354
R	CGG	0.000	6.300	1.700	11.600	0.000	0.000	0.000
R	CGT	0.000	20.300	6.500	4.600	0.000	0.000	0.000
S(Ser)	AGC	46.875	16.000	9.700	19.300	2.930	4.832	2.429
S	AGT	0.000	9.500	14.200	11.900	0.000	0.000	0.000
S	TCA	5.208	7.800	18.800	12.000	0.668	0.277	0.434
S	TCC	10.417	8.900	14.200	11.900	1.170	0.734	0.875
S	TCG	0.000	8.700	8.500	4.400	0.000	0.000	0.000
S	TCT	0.000	8.700	23.500	14.700	0.000	0.000	0.000
T(Thr)	ACA	10.417	8.200	17.800	15.100	1.270	0.585	0.690
T	ACC	36.458	22.800	12.600	19.400	1.599	2.893	1.879
T	ACG	20.833	14.800	7.900	6.100	1.408	2.637	3.415
T	ACT	0.000	9.100	20.300	13.000	0.000	0.000	0.000
V(Val)	GTA	0.000	11.100	11.800	7.200	0.000	0.000	0.000
V	GTC	5.208	15.100	11.600	14.600	0.345	0.449	0.357
V	GTG	0.000	25.500	10.600	28.400	0.000	0.000	0.000
V	GTT	0.000	18.500	22.000	11.000	0.000	0.000	0.000
W(Trp)	TGG	15.625	15.200	10.300	12.700	1.028	1.517	1.230
Y(Tyr)	TAC	10.417	12.100	14.600	15.500	0.861	0.713	0.672
Y	TAT	0.000	16.500	18.900	12.100	0.000	0.000	0.000
*	TAA	5.208	2.000	1.000	0.700	2.604	5.208	7.440
*	TAG	0.000	0.300	0.500	0.600	0.000	0.000	0.000
*	TGA	0.000	1.100	0.700	1.500	0.000	0.000	0.000

4. Discussion

Due to the degeneracy of genetic code, most amino acids are coded by various synonymous codons. And synonymous codons were not used randomly either at equal frequencies in the living organisms[19,

20]. The synonymous codons usage made the different codon usage bias patterns. There were some hypotheses to explain the reasons of codon usage bias, and the typical statement were neutral theory [19] and the 'selection-mutation-drift' model[21, 22]. A lot of research had shown that correlative factors of synonymous codon usage biases were complex, and the major factors were gene expression level, gene length, protein amino acid composition, tRNA abundance, mutation frequency, GC compositions, and so on[20, 23-25].

We had presented a large number analysis of the codon usage patterns in Tianfu goose IFN-α gene and 6 IFN-α genes. According to the analysis of Tianfu goose IFN-α gene, it was revealed that CTC, CAC, GCCM GAC, CGC, CCC, TTC, AGC, CAG, TGC and ACC were Tianfu goose IFN-α gene preferred codons. Meanwhile, Asn, His, Ala, Pro, Arg and Thr were the partial amino acids. The patial amino acids all had 6-fold, 4-fold or 3-fold coding degeneracy. The multiple fold coding made the preferred amino acid composition exist. The corresponding codon was more frequently used than expected while RSCU values greater than 1.0, nevertheless, the corresponding codon was less frequently used than expected while RSCU values smaller than 1.0.

The comparison of codon usage patterns among different animals' IFN-α gene was performed. The results showed that goose and duck IFN-α gene shared the similar patterns, and just displayed slight difference; in addition, chicken IFN-α gene used a distinct pattern. In other words, goose and duck had a much closer relationship than chicken.

Analysis of codon usage bias pattern was academic and actual important in understanding the groundword of molecular biology. Comparison of codon usage bias analysis, the ratio of codon frequency higher than 2 or lower than 0.050 indicated the codon usage preference differed. Comparison of codon usage bias of Tianfu goose IFN-α gene among E.coli, Yeast and Human system(Table 2), the results showed that E.coli was much suitable for the expression of Tianfu goose IFN-α gene than Yeast and Human system.

Research on synonymous codon usage biases was helpful in genetic engineering to rising the yield of target proteins. Our study of codon usage bias has provided an insight into the feature of Tianfu goose IFN-α. We have an initiatory of the function and the characterization of the Tianfu goose IFN-α gene.

References

[1] R. Grantham, C. Gautier, and M. Gouy, "Codon catalog usage and the genome hypothesis," Nucleic acids research, vol. 8, 1980, pp. 197.

[2] A. Lloyd and P. Sharp, "Evolution of codon usage patterns: the extent and nature of divergence between Candida albicans and Saccharomyces cerevisiae," Nucleic acids research, vol. 20, 1992, pp. 5289.

[3] G. Marais and L. Duret, "Synonymous codon usage, accuracy of translation, and gene length in Caenorhabditis elegans," Journal of molecular evolution, vol. 52, 2001, pp. 275-280.

[4] J. Ernst, "Codon usage and gene expression," Trends in Biotechnology, vol. 6, 1988, pp. 196-199.

[5] [K. Sau, S. Gupta, S. Sau *et al.*, "Factors influencing synonymous codon and amino acid usage biases in Mimivirus," Biosystems, vol. 85, 2006, pp. 107-113.

[6] J. Ma, T. Zhou, and W. Gu, "Cluster analysis of the codon use frequency of MHC genes from different species," Biosystems, vol. 65, 2002, pp. 199-207.

[7] J. Vilcek, "The cytokines: an overview," The cytokine handbook, vol. 4, 2003, pp. 3-18.

[8] H. Li, B. Ma, and J. Mi, "Cloning, in vitro expression and bioactivity of goose interferon-[alpha]," Cytokine, vol. 34, 2006, pp. 177-183.

[9] X. Zheng, Y. Liu and S.Chen, Preparation and antiviral effect in vitro of recombinant interferon-alpha of Shitou goose, Chinese Veterinary Science vol. 35, 2008, pp. 29-33.

[10]J. Gutterman, “Cytokine therapeutics: lessons from interferon alpha,” Proceedings of the National Academy of Sciences of the United States of America, vol. 91, 1994, pp. 1198.

[11]E. Borden, “Review: Milstein Award lecture: interferons and cancer: where from here?,” Journal of interferon & cytokine research, vol. 25, 2005, pp. 511-527.

[12]M. Chelbi-Alix, and J. Wietzerbin, “Interferon, a growing cytokine family: 50 years of interferon research,” Biochimie, vol. 89, 2007, pp. 713-718.

[13]P. Sharp, and W. Li, “The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications,” Nucleic acids research, vol. 15, 1987, pp. 1281.

[14]T. Ikemura, “Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes.Differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer RNAs,” Journal of molecular biology, vol. 158, 1982, pp. 573-597.

[15]J. Bennetzen, and B. Hall, “Codon selection in yeast,” Journal of Biological Chemistry, vol. 257, 1982, pp. 3026.

[16]F. Wright, “The 'effective number of codons' used in a gene,” Gene, vol. 87, 1990, pp. 23-29.

[17]T. Ikemura, “Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes* 1,” Journal of molecular biology, vol. 146, 1981, pp. 1-21.

[18]F. Wright, “The'effective number of codons' used in a gene,” Gene, vol. 87, 1990, pp. 23.

[19]Y. Nakamura, T. Gojobori, and T. Ikemura, “Codon usage tabulated from international DNA sequence databases: status for the year 2000,” Nucleic acids research, vol. 28, 2000, pp. 292.

[20]X. Wan, D. Xu, and A. Kleinhofs, “Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes,” BMC Evolutionary Biology, vol. 4, 2004, pp. 19.

[21]P. Sharp, and W. Li, “An evolutionary perspective on synonymous codon usage in unicellular organisms,” Journal of molecular evolution, vol. 24, 1986, pp. 28-38.

[22]M. Bulmer, “Are codon usage patterns in unicellular organisms determined by selection-mutation balance?,” Journal of Evolutionary Biology, vol. 1, 1988, pp. 15-26.

[23]S. Branford, Z. Rudzki, S. Walsh et al., “Detection of BCR-ABL mutations in patients with CML treated with imatinib is virtually always accompanied by clinical resistance, and mutations in the ATP phosphate-binding loop (P-loop) are associated with a poor prognosis,” Blood, vol. 102, 2003, pp. 276.

[24]G. D'Onofrio, T. Ghosh, and G. Bernardi, “The base composition of the genes is correlated with the secondary structures of the encoded proteins,” Gene, vol. 300, 2002, pp. 179-187.

[25]M. Angellotti, S. Bhuiyan, G. Chen et al., “CodonO: codon usage bias analysis within and across genomes,” Nucleic acids research, vol. 35, 2007, pp. W132.